Powers in Sturmian sequences

David Damanik a, Daniel Lenz b

a Department of Mathematics 253–37, California Institute of Technology, Pasadena, CA 91125, USA
b Fakultät für Mathematik, TU Chemnitz, D-09107 Chemnitz, Germany

Received 3 May 2002; accepted 19 December 2002

Abstract

We consider Sturmian sequences and explicitly determine all the integer powers occurring in them. Our approach is purely combinatorial and is based on canonical decompositions of Sturmian sequences and properties of their building blocks. © 2003 Elsevier Science Ltd. All rights reserved.

MSC: 68R15

Keywords: Sturmian sequences; Powers; Combinatorics on words

1. Introduction

In this paper we consider Sturmian sequences and their subword structure. In particular, we are interested in integer powers occurring in them. This paper complements the work [6], which establishes an explicit formula for the index of a Sturmian sequence and hence performs a study of maximal powers. Nevertheless, the general approach of [6] will be the basis of our present study. This approach is combinatorial in nature, and it employs partitions of Sturmian words into canonical words and elementary properties of these words.

Sturmian sequences are aperiodic words of minimal combinatorial complexity. Namely, they are defined over a binary alphabet and they have exactly \( n + 1 \) factors of length \( n \), for every \( n \in \mathbb{N} \). These sequences admit various equivalent definitions and they have been studied from several viewpoints, ranging from computer science to (quasi-)crystallography, resulting in an extensive literature on their structural properties; compare [13].

One important issue is the study of powers in Sturmian sequences. This question has a long history, and among the many significant contributions, we mention [1, 3–7, 9, 14–17]. We will be interested in words \( w \) and integers \( \ell \) such that \( w^\ell \) is a factor of a given Sturmian sequence \( s \). Observe that this set of words and integers depends only on the set \( F(s) \) of...
factors of $s$, rather than the sequence $s$ itself. It is therefore useful to note that for every Sturmian sequence $s$, there is a unique irrational $\alpha \in (0, 1)$ such that

$$F(s) = F(c_\alpha),$$

(1)

where $c_\alpha$ is given by

$$c_\alpha(n) = \chi_{[1-\alpha,1)}(n\alpha \mod 1), \quad n \geq 1,$$

(2)

where $\chi_{[1-\alpha,1)}$ is the characteristic function of the interval $[1 - \alpha, 1)$. The number $\alpha$ is called the slope of $s$. Conversely, every sequence obeying (1) for some irrational $\alpha \in (0, 1)$ is Sturmian, that is, it has minimal complexity among the aperiodic sequences.

The sequence $c_\alpha$ has a hierarchical structure and, in particular, possesses canonical decompositions. This point of view is central to our approach. We shall study the relative positions of powers occurring in $c_\alpha$, corresponding to such a decomposition. This will enable us to find suitable restrictions on the position and the concrete form of such a power. Consequently, we will be able to describe powers in terms of the standard words into which $c_\alpha$ is decomposed.

The organization of this paper is as follows: in the next section, we discuss a few key properties of $c_\alpha$. In particular, we recall the canonical decompositions, which we call $n$-partitions, and some results for them that were obtained earlier. Section 3 determines all the integer powers occurring in $c_\alpha$, for arbitrary irrational $\alpha \in (0, 1)$. Finally, in Section 4, we extend work of Fraenkel and Simpson on the squares occurring in the building blocks of the $n$-partitions, done in [7] for the case $\alpha = (\sqrt{5} - 1)/2$, to general irrational $\alpha \in (0, 1)$.

2. Preliminaries

In this section we discuss some properties of Sturmian sequences we will need in what follows. As general references for Sturmian sequences, we mention [2, 13]. For a general background on combinatorics on words, the reader may also consult [12].

Let us first recall some combinatorial notions. Let $A$ be a finite set, called the alphabet, and denote by $A^*, A^\mathbb{N}$ the set of finite and one-sided infinite words over $A$, respectively. A word $v$ is called a subword (or factor) of some word $u$ if there are words $w_1, w_2$ (possibly empty) such that $u = w_1vw_2$. If a word $w$ can be written as $w = uv$ with words $u, v$, we call $u$ a prefix of $w$ and $v$ a suffix of $w$. Two words $w_1, w_2$ are said to be conjugate, denoted by $w_1 \sim w_2$, if there are words $u, v$ such that $w_1 = uv$ and $w_2 = vu$. Given any word $w$, we denote by $F(w)$ the set of its finite subwords, and by $F(w, n)$ the set of its finite subwords of length $n, n \in \mathbb{N}$. Write $f(w, n)$ for the cardinality of $F(w, n)$; the function $f(w, \cdot) : \mathbb{N} \to \mathbb{N}$ is called the complexity function associated with $w$.

A sequence $s \in \{0, 1\}^\mathbb{N}$, which satisfies

$$f(s, n) = n + 1 \quad \text{for every } n,$$

is called Sturmian. A fundamental theorem of Hedlund and Morse [8] states that a sequence $s$ is ultimately periodic (i.e. there are $n_0, p \in \mathbb{N}$ such that $s_{n+p} = s_n$ for every $n \geq n_0$) if and only if there exists some $N \in \mathbb{N}$ such that $f(s, N) \leq N$. Thus, Sturmian sequences
are exactly the aperiodic sequences of minimal complexity. Clearly, they are defined on an alphabet consisting of two elements, and we can assume that it is given by \( \mathcal{A} = \{0, 1\} \).

For example, every \( c_\alpha \) (as given in (2)) is Sturmian. Conversely, any Sturmian sequence \( s \) is equivalent to some \( c_\alpha \) in the sense that these two sequences have the same factors, \( F(s) = F(c_\alpha) \). Moreover, \( c_\alpha \) has a hierarchical structure which is generated as follows. Recall that every irrational \( \alpha \in (0, 1) \) has a continued fraction expansion

\[
\alpha = [a_1, a_2, a_3, \ldots] = \frac{1}{a_1 + \frac{1}{a_2 + \frac{1}{a_3 + \ddots}}}
\]

with uniquely determined \( a_n \in \mathbb{N} \) (cf. [10]). The associated rational approximants \( p_n/q_n \) are defined by

\[
p_0 = 0, \quad p_1 = 1, \quad p_n = a_n p_{n-1} + p_{n-2},
\]

\[
q_0 = 1, \quad q_1 = a_1, \quad q_n = a_n q_{n-1} + q_{n-2}.
\]

Given the continued fraction expansion of \( \alpha \) as in (3), define the sequence of words \((s_n)_{n \geq -1}\) by

\[
s_{-1} = 1, \quad s_0 = 0, \quad s_1 = s_0^{a_1-1}s_{-1}, \quad s_n = s_{n-1}s_{n-2}, \quad n \geq 2,
\]

so that

\[
|s_n| = q_n \quad \text{for every } n \geq 0.
\]

This sequence of words is called the standard sequence and it is related to \( c_\alpha \) in the following way: it is obvious that for \( n \geq 1 \), \( s_n \) is a prefix of \( s_{n+1} \) and hence there is an obvious meaning to \( \lim_{n \to \infty} s_n \) as an infinite sequence. This sequence is in fact equal to \( c_\alpha \),

\[
c_\alpha = \lim_{n \to \infty} s_n.
\]

We see that \( c_\alpha \) has a hierarchical structure. Indeed, for every \( n \geq 0 \), using (4) and (5), \( c_\alpha \) can be written as a product of blocks of the form \( s_n \) and \( s_{n-1} \). We call this decomposition the \( n \)-partition of \( c_\alpha \) (see [5, 11] as well). Observe that for every \( n \geq 1 \), each \( s_{n-1} \)-block in the \( n \)-partition of \( c_\alpha \) is isolated and \( s_n \)-blocks occur with multiplicity of either \( a_n+1 \) or \( a_n+1+1 \).

Moreover, it follows from (4) that if \( a_1 = 1 \), the combinatorial structures of \( c_\alpha \) and \( c_\alpha \) are the same, where \( \alpha \) has the continued fraction expansion \([a_2 + 1, a_3, a_4, \ldots]\), since this transformation solely results in an exchange of zeros and ones. We may therefore assume without loss of generality that

\[
a_1 \geq 2.
\]

It is often useful to note that the words \( s_n s_{n-1} \) and \( s_{n-1}s_n \) almost coincide. More precisely, from (4) one can deduce the following (cf. Lemma 1 in [3]):

**Lemma 2.1.** For \( n \geq 3 \), we have

\[
s_{n-1}s_n = s_n s_{n-2}^{a_n-1} s_{n-3}s_{n-2}.
\]
Moreover, for \( n \geq 0 \), the words \( s_{n-1}s_n \) and \( s_ns_{n-1} \) agree up to their two rightmost symbols, that is, for \( n \geq 0 \), there is a word \( p_n \) of length \( q_n + q_{n-1} - 2 \) such that
\[
\begin{align*}
s_{n-1}s_n &= \begin{cases} p_n10 & \text{if } n \text{ is even} \\ p_n01 & \text{if } n \text{ is odd} \end{cases} \\
s_n s_{n-1} &= \begin{cases} p_n01 & \text{if } n \text{ is even} \\ p_n10 & \text{if } n \text{ is odd} \end{cases}.
\end{align*}
\]

In the next lemma, we note the well-known primitivity of the words \( s_n \) and \( s_n s_{n-1} \) (see Proposition 2.2.3 and the discussion in Section 2.2.2 of [13]). Recall that a word is primitive if it cannot be written as a nontrivial integer power of another word.

**Lemma 2.2.** For every \( n \) and \( k \), the word \( s_n^k s_{n-1} \) is primitive. In particular, \( s_n \) is primitive and
\[
s_n s_n = us_n v \Rightarrow u = \varepsilon \text{ or } v = \varepsilon.
\]

For later use we also introduce the notation \( R(w) \) to denote the word arising from the word \( w \) by removing its last letter.

The following synchronization lemma (Lemma 3.5 from [6], recall that we are assuming (6)) and its consequences, Lemmas 3.1 and 3.2, will be our main tools in our study of powers in Sturmian sequences.

**Lemma 2.3.** Let \( n \geq 1 \) and consider a factor \( w \) of \( c_\alpha \) with \( q_n \leq |w| < q_{n+1} \). Then the following holds:

(a) There is at most one position in \( s_n \) (resp., \( s_n-1 \)) such that any occurrence of \( w \) in \( c_\alpha \) which starts in some \( s_n \)-block (resp., \( s_n-1 \)-block) of the \( n \)-partition of \( c_\alpha \) has to start at this particular position in \( s_n \) (resp., \( s_n-1 \)).

(b) If \( w \) can start at position \( k \) in \( s_n \) and at position \( l \) in \( s_{n-1} \) (\( k \) and \( l \) are unique by (a)), then we have \( k = l \).

### 3. Squares, cubes, and higher powers in a Sturmian sequence

In this section, we identify all the squares, cubes, and other integer powers that occur in a given Sturmian sequence.

Fix \( m, \ell \in \mathbb{N} \), where \( \ell \geq 2 \), and \( \alpha \in (0, 1) \) irrational, and define the following sets and numbers:
\[
P(m; \ell, \alpha) = \{ w \in [0, 1]^\ast : |w| = m, \ w^\ell \in F(c_\alpha) \}
\]
and
\[
p(m; \ell, \alpha) = |P(m; \ell, \alpha)|,
\]
where \(| \cdot | \) denotes cardinality. For example, \( p(5; 2, \alpha) \) is the number of squares of length 10 (=5 \cdot 2) in \( c_\alpha \). We clearly have the following inequality:
\[
p(m; \ell_1, \alpha) \geq p(m; \ell_2, \alpha) \quad \text{if } \ell_1 \leq \ell_2.
\]

Our first result puts a considerable restriction on the possible values for the length \( m \) such that there is a nontrivial power of a word of length \( m \) in \( c_\alpha \):
Theorem 1. Let \( n \in \mathbb{N} \) and \( q_n \leq m < q_{n+1} \).
Suppose that
\[
m \notin \{ kq_n : 1 \leq k \leq a_{n+1} \} \cup \{ kq_n + q_{n-1} : 1 \leq k \leq a_{n+1} - 1 \}.
\] (11)
Then, for every \( \ell \geq 2 \), we have \( p(m; \ell, \alpha) = 0 \).

Proof. By (10), we only have to show that for \( m \) satisfying (11), we have
\[
p(m; 2, \alpha) = 0.
\] (12)
Assume that (12) fails for some \( m \) and let \( u \) be a word of length \( m \) with \( q_n \leq m < q_{n+1} \)
such that \( u^2 \) is a factor of \( c_{\alpha} \). Consider an occurrence of \( u^2 \) in \( c_{\alpha} \). In the following we will often write \( u^2 \) as \( u(1)u(2) \) to be able to refer to each one of the two occurring blocks. Now, \( u(1) \) starts in some \( s_n \)-block or in some \( s_{n-1} \)-block at position \( l \) and, by Lemma 2.3, \( u(2) \)
also starts in some block of the \( n \)-partition of \( c_{\alpha} \) at position \( l \). Since \( |u| < q_{n+1} \) and two successive \( s_{n-1} \)-blocks in the \( n \)-partition are separated by at least \( a_{n+1} \) blocks of \( s_n \)-type, we get \( |u| = kq_n + jq_{n-1} \) for \( 1 \leq k \leq a_{n+1} \) and \( j = 0 \) or 1. Since \( |u| < q_{n+1} \) and \( q_{n+1} = a_{n+1}q_n + q_{n-1} \), we see that the case \( k = a_{n+1} \) and \( j = 1 \) is impossible. We conclude that \( m \) does not satisfy (11). \( \square \)

Lemma 3.1. Let \( n \geq 1 \) and let \( a \) be the last letter of \( s_{n-1} \). Then the word \( as_n \) occurs at
position \( j \) in \( c_{\alpha} \) if and only if the \( n \)-partition of \( c_{\alpha} \) contains an \( s_n \) starting at \( j \) and an \( s_{n-1} \)
ending at \( j - 1 \). Moreover, \( aR(s_n) \) occurs exactly at those places where \( as_n \) occurs.

Proof. By Lemma 2.3, the word \( s_n \) can only occur at the starting positions of \( s_n \) and \( s_{n-1} \)
in the \( n \)-partition. Now, the first statement follows, as the \( s_{n-1} \) occur with power one and
the last letter of \( s_{n-1} \) is different from the last letter of \( s_n \). As for the last statement, note that \( |aR(s_n)| = q_n \). Obviously, \( aR(s_n) \) occurs at every last position of an \( s_{n-1} \) of the \( n \)-partition.
By Lemma 2.3 again, the only other position where \( aR(s_n) \) may occur is the \( |s_{n-1}| \)-position of an \( s_n \). But then it would end with the penultimate letter of \( s_{n-1} \), as to the
right of an \( s_n \) the \( n \)-partition there appears the word \( s_{n-1} \). This, however, is impossible as the penultimate letter of \( s_{n-1} \) is different from the penultimate letter of \( s_n \). \( \square \)

Lemma 3.2. Let \( n \in \mathbb{N} \) be given. Let \( a \) be the last letter of \( s_{n-1} \). Let \( u^2 = u(1)u(2) \)
be an occurrence of \( u^2 \) in \( c_{\alpha} \), where \( q_n \leq |u| < q_{n+1} \). Then, \( u \) does not contain \( aR(s_n) \).

Proof. Assume that \( aR(s_n) \) is a subword of \( u \). Then it occurs in both \( u(1) \) and \( u(2) \) at
the same position. However, by Lemma 3.1, different occurrences of \( aR(s_n) \) correspond to
different occurrences of \( s_{n-1} \)-blocks in the \( n \)-partition and thus have distance at least \( q_{n+1} \).
This gives a contradiction as \( |u| < q_{n+1} \). \( \square \)

Equipped with these lemmas, we study squares and higher powers of words of length \( m \)
occurring in \( c_{\alpha} \), where \( m \) does not belong to the set excluded by Theorem 1. To this end,
we introduce the cyclic shift operator \( C : \{0,1\}^* \to \{0,1\}^* \) which is defined by
\[
C(x_1x_2 \ldots x_m) = x_2 \ldots x_m x_1,
\]
where \( x_j \in \{0,1\} \) for \( j = 1, \ldots, m \). In particular, this yields for \( 1 \leq j \leq m - 1 \),
\[
C^j(x_1x_2 \ldots x_m) = x_{j+1} \ldots x_m x_1 \ldots x_j.
\]
Observe that two words $u$, $v$ are conjugate if and only if there is $j$ such that $u$ can be written as $u = c^j(v)$.

We first discuss squares of words of length $m < a_1 = q_1$.

**Theorem 2.** For $1 \leq k \leq a_1 - 1$, we have

$$p(k; 2, \alpha) = \begin{cases} 1 & \text{if } k \leq a_1/2 \\ 0 & \text{if } k > a_1/2, \end{cases}$$

where for $k \leq a_1/2$, the unique square is given by $(0^k)^2$.

**Proof.** Consider a factor $u$ of $c_a$ with $k = |u| \leq a_1 - 1$. Then $u$ is either given by $0^k$ or it is conjugate to $0^{k-1}1$. Since $0$ occurs in the $0$-partition of $c_a$ with multiplicity $a_1 - 1$ or $a_1$, it is immediate that there are no squares of words conjugate to $0^{k-1}1$. For words of the form $0^k$, on the other hand, it is clear by the same argument that there is a square if and only if $2k \leq a_1$, and in this case, there is exactly one square of a word of a given length. This gives (13). □

We next consider squares of words of length $m \geq q_1$. We start with some preliminary considerations.

**Lemma 3.3.** Let $u^2 = u^{(1)}u^{(2)}$ be an occurrence of $u^2$ in $c_a$, where $q_n \leq |u| < q_{n+1}$.

(a) If $|u| = q_n$, then $u^{(1)}$ begins in an $s_n$-block of the $n$-partition. Moreover, $u^{(1)}u^{(2)}$ is contained in $s_n^{a_n+1+2}R^2(s_n)$, where $l = a_n - q_n + 2$.

(b) If $|u| = q_n + q_n - 1$, then $u^{(1)}$ contains an $s_{n-1}$-block of the $n$-partition. Moreover, $u^*$ is contained in $s_n^{a_n+1}s_{n-1}^{a_{n-1}}s_{n-1}^{a_{n-1}}s_{n-1}$, which in turn is contained in $s_n^{a_n+1}s_{n-1}^{a_{n-1}}s_{n-1}^{a_{n-1}}s_{n-1}$.

**Proof.** Let $a$ be the last letter of $s_{n-1}$.

(a) The first statement of (a) is a direct consequence of $|u| = q_n$, with $1 \leq k \leq a_{n+1}$ and *Lemma 2.3* (see the proof of *Theorem 1* for related arguments). As for the second statement, recall that the $s_n$-block in which $u^{(1)}$ starts is followed by $u = s_n^a s_{n-1}^b s_n$, where $r \in \mathbb{N}_0$ with $0 \leq r \leq a_{n+1}$. As $R^2(s_n) = s_n^{a_n+2}R(s_n)$ by (8), the second statement can only be wrong if $u^{(2)}$ contains $R(s_{n-1}^{a_{n-1}})$, which in turn contains $aR(s_n)$. But this is impossible by *Lemma 3.2*.

(b) From *Lemma 2.3* (see the proof of *Theorem 1* for related arguments as well), we infer that $u^{(1)}$ either starts in an $s_{n-1}$-block of the $n$-partition or contains such a block. However, in the first case, by $|u| \leq q_n + q_{n-1}$, the word $u$ would contain $aR(s_n)$ as $s_{n-1}$ is followed by $s_n$. This is a contradiction to *Lemma 3.2*, and only the second case can occur. Thus, $u^{(1)}$ contains a block $s = s_{n-1}$ of the $n$-partition. By *Lemma 2.3*, it ends before the last two letters of the $s_n$-block to the right of $s$. As to the left of $s$, there are $a_{n+1}$ blocks of the form $s_n$, we infer from $|u| = q_n + q_{n-1}$ with $1 \leq k \leq a_{n+1}$ that there exists $j \in \mathbb{N}_0$, with $0 \leq j < q_n - 2$ such that $u$ starts at position $j$ in $s_n^{a_n+1}s_{n-1}$, $s_{n-1}^{a_{n-1}}$. But this means that $u^2$ is a subword of $s_n^{a_n+1}s_{n-1}^b s_{n-1}^{a_{n-1}}R^2(s_n)$. Furthermore, invoking (8), it is straightforward to see that $s_n^{a_n+1}s_{n-1}^b s_{n-1}^{a_{n-1}}R^2(s_n)$ is contained in $s_n^{a_n+1}s_{n-1}^b s_{n-1}^{a_{n-1}}s_{n-1}$ for every $k$ with $1 \leq k \leq a_{n+1} - 1$. □

**Theorem 3.** Let $n \in \mathbb{N}$. 


Let \( n \) be given. Let \( a \) be the last letter of \( s_n \). Then \( \alpha \) is a word with \( \alpha \) of the form \( \mathcal{C}^{1}(s_{n}^{h} s_{n-1} \alpha) \), with \( 0 \leq |\alpha| = k q_n + q_n - 1 < q_{n-1} - 2 \), \( q_n \), and \( q_{n-1} \) are even and \( k = 1 + a_{n+1}/2 \). More precisely, when \( 1 \leq k < 1 + a_{n+1}/2 \), we have

\[
\mathcal{P}(k q_n; 2, \alpha) = \{ \mathcal{C} \mathcal{C}^{1}(s_{n}^{h} s_{n-1} \alpha) : 0 \leq j < q_{n-1} - 2 \},
\]

and when \( a_{n+1} \) is even and \( k = 1 + a_{n+1}/2 \), we have

\[
\mathcal{P}(k q_n; 2, \alpha) = \{ \mathcal{C} \mathcal{C}^{1}(s_{n}^{h} s_{n-1} \alpha) : 0 \leq j < q_{n-1} - 2 \}.
\]

We have for \( k q_n + q_n - 1 \), \( \alpha \) = \( q_{n-1} - 2 \).

More precisely,

\[
\mathcal{P}(k q_n + q_n - 1; 2, \alpha) = \{ \mathcal{C} \mathcal{C}^{1}(s_{n}^{h} s_{n-1} \alpha) : 0 \leq j < q_{n-1} - 2 \}.
\]

**Proof.** Essentially, we prove (15), (16), and (18). Lemma 2.2, along with (9), implies that the elements in the respective sets are mutually distinct, which in turn yields the claimed statements (14) and (17) for \( p(m; 2, \alpha) \).

(a) As discussed earlier, \( s_{n}^{a_{n+1}+1} s_{n-1} s_{n} \) is contained in \( c_{\alpha} \). In particular, \( w = R^{2} (s_{n}^{a_{n+1}+1} s_{n-1} s_{n}) \) is contained in \( c_{\alpha} \). As by (8), \( w = s_{n}^{a_{n+1}+2} R^{1}(s_{n}) \), where \( l = q_{n} - q_{n-1} + 2 \), it suffices by Lemma 3.3 (a) to find all squares \( u^{2} \) of words \( u \), with \( |u| = k q_n \), that appear in \( w \). Of course it suffices to consider occurrences of \( u^{2} \) starting in the first \( s_{n} \)-block of \( w \). Now, (a) is immediate.

(b) Choose \( k \in \mathbb{N} \) with \( 1 \leq k \leq a_{n+1} - 1 \). Clearly, the word \( s_{n}^{a_{n+1}+1} s_{n-1} s_{n}^{a_{n+1}+1} s_{n-1} s_{n} \) is a factor of \( c_{\alpha} \). Thus, every square of a word \( u \) of the form \( u = \mathcal{C} \mathcal{C}^{1}(s_{n}^{h} s_{n-1} \alpha) \), with \( 0 \leq j < q_{n-1} - 2 \), occurs in \( c_{\alpha} \). On the other hand, if \( u \) is a word with \( |u| = k q_n + q_n - 1 \), for which \( u^{2} \) occurs in \( c_{\alpha} \), then \( u^{2} \) is a subword of \( w \) by Lemma 3.3. As \( |w| = 2|u| + q_n - 2 \), we infer that \( u \) must be of the form \( u = \mathcal{C} \mathcal{C}^{1}(s_{n}^{h} s_{n-1} \alpha) \) with \( 0 \leq j \leq q_{n-1} - 2 \).

Now that we understand the structure and the occurrences of squares in \( c_{\alpha} \), we turn to higher integer powers. Their study is substantially facilitated by the theorem above since every power \( w^{p} \) of some word of length \( |w| = m \) with \( p \geq 3 \) contains \( p - 1 \) occurrences of the square \( w^{2} \), so Theorem 3 helps us in identifying the possible words \( w \) and the possible consecutive occurrences of \( w^{2} \). In fact, the proof of the following theorem will be an immediate consequence of Theorem 3 and the following lemma.

**Lemma 3.4.** Let \( n \in \mathbb{N} \) be given. Let \( \alpha \) be the last letter of \( s_{n-1} \). Let \( u^{3} = \mathcal{C} \mathcal{C}^{1}(s_{n}^{h} s_{n-1} \alpha) \) be an occurrence of \( u^{3} \), where \( q_n \leq |\alpha| < q_{n+1} \) in \( c_{\alpha} \). Then \( u^{3} \) does not contain \( a R(s_{n}) \).

**Proof.** Assume the contrary. By Lemma 3.2, \( a R(s_{n}) \) cannot be contained in \( u^{(3)} \). Thus, every occurrence of \( a R(s_{n}) \) must start in \( u^{(1)} \) or in \( u^{(2)} \). As both \( u^{(1)} \) and \( u^{(2)} \) are followed by a copy of \( u \), there exists a \( j \in \mathbb{N}_{0} \) such that \( a R(s_{n}) \) starts at the \( j \)th position in both \( u^{(1)} \)
and \( u^{(2)} \). As occurrences of \( a R(s_n) \) have distance at least \( q_{n+1} \) by Lemma 3.1, we obtain the contradiction \( |u| \geq q_{n+1} \). \( \square \)

**Theorem 4.** Let \( n, p \in \mathbb{N}, p \geq 3 \).

(a) We have for \( 1 \leq k \leq a_{n+1} \),

\[
p(kq_n; p, \alpha) = \begin{cases} q_n & \text{if } 1 \leq k < (a_{n+1} + 2)/p \\ q_{n-1} - 1 & \text{if } k = (a_{n+1} + 2)/p \\ 0 & \text{if } (a_{n+1} + 2)/p < k \leq a_{n+1} .
\end{cases}
\]  

(b) We have for \( 1 \leq k \leq a_{n+1} - 1 \),

\[
p(kq_n + q_{n-1}; p, \alpha) = 0 .
\]  

**Proof.** (a) This part is similar to the proof of part (a) of Theorem 3. Consider an occurrence of \( u^p = u^{(1)} \ldots u^{(p)} \) in \( c_\alpha \), where \( |u| = kq_n \) with \( 1 \leq k \leq a_{n+1} \) and \( p \geq 3 \). By Lemma 3.3 (b), \( u^{(1)}u^{(2)} \) begins in an \( s_n \)-block of the \( n \)-partition.

By Lemma 3.4, \( u^p \) cannot contain \( a R(s_n) \). Thus, as in the proof of Theorem 3, we infer that \( u^p \) is contained in \( w = b_n^{a_{n+1}+2} R^l(s_n), \) where \( l = q_n - q_{n-1} + 2 \). Now, the proof of (a) can be finished along the same lines as the proof of Theorem 3.

(b) Let \( u \) be given with \( |u| = kq_n + q_{n-1} \) with \( 1 \leq k \leq a_{n+1} - 1 \). Assume that \( u^3 = u^{(1)}u^{(2)}u^{(3)} \) occurs in \( c_\alpha \). Then \( u^{(1)} \) contains an \( s_{n-1} \)-block of the \( n \)-partition by Lemma 3.3 (c). Then \( u^3 \) contains \( a R(s_n) \), where \( a \) is the last letter of \( s_{n-1} \), and a contradiction to Lemma 3.4 follows. \( \square \)

4. The functions \( D \) and \( R \) of Fraenkel and Simpson

In this section we extend results of Fraenkel and Simpson [7] who studied the squares contained in the word \( s_n \), for \( n \in \mathbb{N} \), in the Fibonacci case, that is, \( \alpha = (\sqrt{5} - 1)/2 \) and \( a_n = 1 \) for every \( n \in \mathbb{N} \), to the case of general irrational \( \alpha \in (0, 1) \).

Given some irrational \( \alpha \in (0, 1) \), we define for \( n \in \mathbb{N} \), the following functions:

\[
D : \mathbb{N} \to \mathbb{N}_0, D(n) = \text{the number of distinct squares in } s_n ,
\]

\[
R : \mathbb{N} \to \mathbb{N}_0, R(n) = \text{the number of repeated squares in } s_n .
\]

Our goal in this section is to study these functions with the help of the results of the previous section. More precisely, we show how the results of Section 3 allow one to provide an explicit expression for \( D(n) \), while for \( R(n) \), we merely note that a similar analysis can be performed.

Assume now for a while that (6) holds. As we noted above, this does not lead to any real restriction. Let us introduce some auxiliary quantities. For \( n \in \mathbb{N} \) and \( 0 \leq m \leq n - 1 \),
let $d(n, m)$ denote the number of distinct squares occurring in $s_n$ which are naturally associated with the $m$-partition of $s_n$. Here, the $m$-partition of $s_n$ is the canonical decomposition of $s_n$ into blocks of the form $s_m$ and $s_{m-1}$ which follows from (4), just as in the case of $c_\alpha$. Moreover, $u^\alpha$ is said to be naturally associated with the $m$-partition of $s_n$ if $q_m \leq |u| < q_{m+1}$. Note that such a word $u$ satisfies $|u| = kq_m$ with $1 \leq k \leq a_{m+1}$ or $|u| = kq_m + q_{m-1}$ with $1 \leq k \leq a_{m+1} - 1$ by Theorem 1.

The numbers $d(n, m)$ are, in principle, easy to compute, given the results of the previous section, but we have to distinguish many cases and hence split up their investigation into a series of lemmas. We let $d_m$ denote the number of distinct squares occurring in $c_\alpha$ which are naturally associated with the $m$-partition of $c_\alpha$, that is, for $m \geq 1$, we have

$$d(n, m) = d_1(m) + d_2(m),$$

where

$$d_1(m) = \sum_{k=1}^{a_{m+1}} p(kq_m; 2, \alpha) \quad \text{and} \quad d_2(m) = \sum_{k=1}^{a_{m+1}-1} p(kq_m + q_{m-1}; 2, \alpha),$$

while for $m = 0$, we have by Theorem 2

$$d(0) = \left\lfloor \frac{a_1}{2} \right\rfloor.$$

In our first lemma, we describe all the cases where $s_n$ contains all the squares on a given level that occur in $c_\alpha$. The subsequent lemmas consider the cases where some of the squares that occur in $c_\alpha$ do not occur in $s_n$.

**Lemma 4.1.** We have for $m \leq n - 4$, or for $m = n - 3$ and $a_n \geq 2$,

$$d(n, m) = d(m). \tag{23}$$

**Proof.** We clearly have

$$d(n, m) \leq d(m), \tag{24}$$

so that we only have to prove $d(n, m) \geq d(m)$ in order to establish (23). To this end, it will be sufficient to show that the finite words containing all the squares in $c_\alpha$ on level $m$ that were found in the previous section are all contained in $s_n$. More precisely, by Lemma 3.3, we have to show that the words

$$R^2(s_m^{a_n+1} s_{m-1}^{a_n+1}) \quad \text{and} \quad s_m^{a_n+1} s_{m-1}^{a_n+1}$$

are subwords of $s_n$. Expanding $s_n$ to level $m$, using (4), one may verify that under the assumption of the lemma, this claim holds. For example, in the case where $m = n - 3$ and $a_n \geq 2$, we obtain

$$s_n = ((s_{n-3}^{a_n-2} s_{n-4})^{a_n-1} s_{n-3})^{a_n} s_{n-3}^{a_n-2} s_{n-4}.$$  

Since $a_n \geq 2$, $s_n$ contains

$$(s_{n-3}^{a_n-2} s_{n-4})^{a_n-1} s_{n-3} (s_{n-3}^{a_n-2} s_{n-4})^{a_n-1} s_{n-3}.$$
and therefore, by (8), it also contains
\[ R^2(s_{n-3}^{a_n-2+2} s_{n-4}) \quad \text{and} \quad s_{n-3}^{a_n-2} s_{n-4} s_{n-3}^{a_n-2+1} s_{n-4}. \]

In the next lemma, we study the remaining case, \( a_n = 1 \), for \( m = n - 3 \).

**Lemma 4.2.** When \( a_n = 1 \), we have
\[
d(n, n - 3) = \begin{cases} 
\frac{d(n - 3) - q_{n-4} + 1}{2} & \text{if } a_{n-2} \text{ is even} \\
\frac{d(n - 3) - q_{n-3} + q_{n-4} + 1}{2} & \text{if } a_{n-2} \text{ is odd}.
\end{cases}
\] (25)

**Proof.** We expand \( s_n \) to level \( n - 3 \), using (4),
\[ s_n = (s_{n-3}^{a_{n-2}} s_{n-4}) s_{n-3}^{a_{n-2}} s_{n-3}^{a_{n-2}} s_{n-4}. \]

Thus, \( s_n \) contains \( s_{n-3}^{a_{n-2}} s_{n-4} s_{n-3}^{a_{n-2}} s_{n-4} \) and hence it contains all \( d_2(n - 3) \) squares of words conjugate to \( s_{n-3}^{a_{n-2}+1} s_{n-4} \) for \( 1 \leq k \leq a_{n-2} - 1 \).

On the other hand, \( s_n \) does not contain the word \( R^2(s_{n-3}^{a_{n-2}+2} s_{n-4}) \), as, by (8) and Lemma 2.3, \( a_{n-2} + 2 \) copies of \( s_{n-2} \) can only occur at those places, where there are \( a_{n-2} + 1 \) copies of \( s_{n-2} \) in the \( n \)-partition. The maximal power of \( s_{n-3} \) occurring in \( s_n \) is then rather the word \( s_{n-3}^{a_{n-2}+1} s_{n-4} \).

We first consider the case where \( a_{n-2} \) is even. Then there are \( q_{n-4} - 1 \) squares in \( R^2(s_{n-3}^{a_{n-2}+2} s_{n-4}) \) that do not occur in \( s_{n-3}^{a_{n-2}+1} s_{n-4} \) (the ones of maximal length). We therefore get
\[
d(n, n - 3) = (d_1(n - 3) - (q_{n-4} - 1)) + d_2(n - 3) = d(n - 3) - q_{n-4} + 1.
\]

Consider now the case where \( a_{n-2} \) is odd. In this case, the \( s_{n-4} \)-block in \( R^2(s_{n-3}^{a_{n-2}+2} s_{n-4}) \) is not needed and we find all the relevant squares already in \( s_{n-3}^{a_{n-2}+2} \). Comparing this with \( s_{n-3}^{a_{n-2}+1} s_{n-4} \), we see that there are \( q_{n-3} - q_{n-4} - 1 \) squares in \( s_{n-3}^{a_{n-2}+2} \) that are not contained in \( s_{n-3}^{a_{n-2}+1} s_{n-4} \). That is, we have
\[
d(n, n - 3) = (d_1(n - 3) - (q_{n-3} - q_{n-4} - 1)) + d_2(n - 3)
= d(n - 3) - q_{n-3} + q_{n-4} + 1,
\]
completing the proof. \( \square \)

Next, we study the case \( m = n - 2 \):

**Lemma 4.3.** (a) When \( a_n = 1 \), we have
\[
d(n, n - 2) = \begin{cases} 
\frac{d_1(n - 2) - q_{n-3} + 1}{2} & \text{if } a_{n-1} \text{ is even} \\
\frac{d_1(n - 2) - q_{n-2} + q_{n-3} - 1}{2} & \text{if } a_{n-1} \text{ is odd}.
\end{cases}
\] (26)

(b) When \( a_n \geq 2 \), we have
\[
d(n, n - 2) = \begin{cases} 
\frac{d(n - 2) - q_{n-3} + 1}{2} & \text{if } a_{n-1} \text{ is even} \\
\frac{d(n - 2) - q_{n-2} + q_{n-3} - 1}{2} & \text{if } a_{n-1} \text{ is odd}.
\end{cases}
\] (27)

**Proof.** (a) When \( a_n = 1 \), we have
\[ s_n = s_{n-2}^{a_{n-1}} s_{n-3} s_{n-2}. \]
As above, we have to compare this with

\[ R^2(s_{n-2}^{a_n-1+2}s_{n-3}) \quad \text{and} \quad s_{n-2}^{a_n-1}s_{n-3}s_{n-2}^{a_n-1}. \]

Clearly, the \( d_2(n-2) \) squares of words conjugate to \( s_{n-2}^{a_n-1}s_{n-3} \) do not occur in \( s_n \) and hence we have, in this case, \( d(n, n-2) \leq d_1(n-2) \). For the squares coming from \( R^2(s_{n-2}^{a_n-1+2}s_{n-3}) \), when compared with \( R^2(s_n) = R^2(s_{n-2}^{a_n-1}s_{n-3}s_{n-2}) = R^2(s_{n-2}^{a_n-1+1}s_{n-3}) \), we see that for \( a_n-1 \) even, we lose \( q_{n-2} - (q_{n-3} - 1) \) squares, while for \( a_n-1 \) odd, we lose \( q_{n-2} - (q_{n-3} - 1) \) squares, leading to (26) (see the proof of the preceding lemma for similar arguments).

(b) When \( a_n \geq 2 \), we have

\[ s_n = (s_{n-2}^{a_n-1}s_{n-3})^{a_n}s_{n-2}, \]

which contains

\[ s_{n-2}^{a_n-1}s_{n-3}s_{n-2}^{a_n-1}s_{n-3}s_{n-2}. \]

Hence it contains all \( d_2(n-2) \) squares of words conjugate to \( s_{n-2}^{a_n-1}s_{n-3} \).

For squares coming from \( R^2(s_{n-2}^{a_n-1+2}s_{n-3}) \), we again have to compare this with \( R^2(s_{n-2}^{a_n-1+1}s_{n-3}) \), and the same arguments as in part (a) lead to (27). □

Finally, we consider squares in \( s_n \) on the level \( m = n - 1 \):

**Lemma 4.4.** We have

\[ d(n, n-1) = \begin{cases} 
    d_1(n-1) - q_{n-1} + 2 & \text{if } a_n \text{ is even} \\
    d_1(n-1) - q_{n-1} & \text{if } a_n \text{ is odd.}
\end{cases} \quad (28) \]

**Proof.** We have

\[ s_n = s_{n-1}^{a_n}s_{n-2}. \]

All the squares in \( c_\alpha \) on the level of the \((n-1)\)-partition are contained in the words

\[ R^2(s_{n-1}^{a_n+2}s_{n-2}) \quad \text{and} \quad s_{n-1}^{a_n}s_{n-2}s_{n-1}^{a_n}s_{n-2}. \]

Since the \((n-1)\)-partition of \( s_n \) has only one \( s_{n-2} \)-block, we have no squares of words conjugate to \( s_{n-1}^{a_n}s_{n-2} \) occurring in \( s_n \), so we get \( d(n, n-1) \leq d_1(n-1) \).

Consider first the case where \( a_n \) is even. All the \( d_1(n-1) \) squares in \( c_\alpha \) are contained in \( R^2(s_{n-1}^{a_n+2}s_{n-2}) \). This word does not occur in \( s_n \), so its role is played by \( s_{n-1}^{a_n}s_{n-2} \). We see that we lose \( q_{n-1} \) squares since two \( s_{n-1} \)-blocks are lacking, but we gain two additional squares of maximal length since the last two symbols of \( s_{n-2} \) can be used. In other words, we lose \( q_{n-2} - 2 \) squares of maximal length as well as \( q_{n-1} - q_{n-2} \) squares of length \( q_{n-1}(a_n)/2 \). Thus, we obtain \( d(n, n-1) = d_1(n-1) - q_{n-1} + 2 \) in this case.

Consider now the case where \( a_n \) is odd. Then the \( s_{n-2} \)-block is not needed for squares in either \( c_\alpha \) and \( s_n \), and we just lose \( q_{n-1} \) squares due to the two less \( s_{n-1} \)-blocks. We get \( d(n, n-1) = d_1(n-1) - q_{n-1} \) in this case.

This shows (28) and concludes the proof. □
Now that we have found explicit expressions for the numbers \( d(n, m) \), we can conclude our study of the function \( D \) introduced above by noting the following:

**Theorem 5.** Let \( n \in \mathbb{N} \). Then we have

\[
D(n) = \sum_{m=0}^{n-1} d(n, m). \tag{29}
\]

**Proof.** The inequality “\( \geq \)” is obvious and the inequality “\( \leq \)” follows as all squares appearing in \( s_n \) are in fact counted by some \( d(n, m) \) with \( m \leq n-1 \) (see proof of Theorem 1 as well). \( \square \)

We consider a few examples. First, let us recover the result of Fraenkel and Simpson [7]. Namely, we consider the golden mean case, \( \alpha = (\sqrt{5} - 1)/2 \), where \( a_n = 1 \) for every \( n \in \mathbb{N} \). In this case, the lengths of the \( s_n \)’s just follow a Fibonacci sequence \( F_n \) defined by

\[
F_0 = 1, \quad F_1 = 1, \quad F_{n+1} = F_n + F_{n-1} \quad \text{for } n \geq 1.
\]

Since we were working under the normalization condition (6) above, we consider \( c_{\hat{\alpha}} \), where \( \hat{\alpha} \) has continued fraction coefficients

\[
a_1 = 2, \quad a_n = 1 \quad \text{for } n \geq 2.
\]

In particular, we obtain for the \( q_n \)’s associated with \( \hat{\alpha} \),

\[
q_n = F_{n+1} \quad \text{for every } n \geq 1. \tag{30}
\]

Let us find the numbers \( d(m) \) associated with \( c_{\hat{\alpha}} \). As we noted before, \( c_{\hat{\alpha}} \) results from \( c_{\hat{\alpha}} \) by an interchange of zeros and ones. However, on the level of the \( s_n \)’s, we have to shift the index by one due to the length condition (30). Hence, we have

\[
D_{\alpha}(n) = D_{\hat{\alpha}}(n-1) \quad \text{for every } n \in \mathbb{N}.
\]

From (14), we get \( p(q_n; 2, \hat{\alpha}) = q_n \) and hence

\[
d(m) = d_1(m) = q_m \quad \text{for every } m \in \mathbb{N}.
\]

**Theorem 2** gives

\[
d(0) = 1.
\]

Therefore, fixing some \( n \geq 5 \), (23) yields

\[
d(n-1, m) = q_m \quad \text{for every } 1 \leq m \leq n - 5
\]

and

\[
d(n-1, 0) = 1.
\]

From (25), we obtain

\[
d(n-1, n-4) = q_{n-5} + 1
\]

and (26) yields

\[
d(n-1, n-3) = q_{n-4} - 1.
\]
Finally, (28) gives
\[ d(n - 1, n - 2) = 0 \]
and we can summarize our findings as follows (cf. Theorem 1 in [7]):

**Corollary 1.** Let \( \alpha = (\sqrt{5} - 1)/2 \) and \( n \geq 5 \). Then
\[ D(n) = 2(F_{n-2} - 1). \]

**Proof.** By what we have seen above, and (30) in particular, we have
\[
D(n) = 1 + \sum_{m=1}^{n-5} q_m + (q_{n-5} + q_{n-4}) \\
= \sum_{m=1}^{n-4} F_m + (F_{n-4} + F_{n-3}) \\
= (F_{n-4} + F_{n-5} + \cdots + F_3 + F_2 + F_1) + (F_2 - 2) + (F_{n-4} + F_{n-3}) \\
= F_{n-4} + F_{n-5} + \cdots + F_3 + F_2 + F_1 + 2 + F_{n-2} \\
= F_{n-4} + F_{n-5} + \cdots + F_3 + F_2 + F_1 - 2 + F_{n-2} \\
\vdots \\
= F_{n-4} + F_{n-3} - 2 + F_{n-2} \\
= F_{n-2} - 2 + F_{n-2} \\
= 2(F_{n-2} - 1). \quad \square
\]

Our next example is given by the silver mean, \( \alpha = \sqrt{2} - 1 \), where \( a_n = 2 \) for every \( n \in \mathbb{N} \). From (14), we get \( p(q_n; 2, \alpha) = q_n \) and \( p(2q_n; 2, \alpha) = q_{n-1} - 1 \), and from (17), we get \( p(q_n + q_{n-1}; 2, \alpha) = q_n + q_{n-1} - 1 \). Hence
\[ d(m) = 2q_m + 2q_{m-1} - 2 \quad \text{for every } m \in \mathbb{N}. \]

**Theorem 2** gives
\[ d(0) = 1. \]

Therefore, fixing some \( n \geq 4 \), (23) yields
\[ d(n, m) = 2q_m + 2q_{m-1} - 2 \quad \text{for every } 1 \leq m \leq n - 3 \]
and
\[ d(n, 0) = 1. \]

From (27), we obtain
\[ d(n, n - 2) = 2q_{n-2} + q_{n-3} - 1. \]

Finally, (28) gives
\[ d(n, n - 1) = q_{n-2} + 1 \]
and hence, by **Theorem 5**, we find:
Corollary 2. Let \( \alpha = \sqrt{2} - 1 \) and \( n \geq 4 \). Then
\[
D(n) = 1 + 3q_{n-2} + q_{n-3} + \sum_{m=1}^{n-3} 2q_m + 2q_{m-1} - 2.
\]

It is clear that a completely analogous analysis is possible, which yields an explicit formula for the function \( R \) in the general Sturmian case (Fraenkel and Simpson give such a formula in the case of the golden mean). We do not work this out in detail, and we just remark that all the necessary tools have been established, so that the task is rather straightforward (albeit somewhat cumbersome).

Acknowledgements

D.D. was supported in part by the NSF grant DMS-0227289. D.L. was partially supported by the DFG in the priority program Quasikristalle.

References